

Background

Definition

The overall BLS definition of green jobs includes two components -- output and process -- that will be used for two separate BLS products. This release uses the “output” definition.

Under the BLS “output” approach, green jobs are *Jobs in businesses that produce goods and provide services that benefit the environment or conserve natural resources*. These goods and services are sold to customers, and include research and development, installation, and maintenance services. Green goods and services fall into one or more of five groups:

1. *Energy from renewable sources.*
2. *Energy efficiency.*
3. *Pollution reduction and removal, greenhouse gas reduction, and recycling and reuse.*
4. *Natural resources conservation.*
5. *Environmental compliance, education and training, and public awareness.*

To implement the output approach, BLS collects data on jobs associated with producing green goods and services through a sample survey of establishments identified as potentially producing such products and services based on their 2007 North American Industry Classification System (NAICS) classification. The purpose of the Green Goods and Services (GGS) survey is to identify whether the establishment is producing any green goods and services and, if so, to measure the number of associated jobs in the establishment.

The GGS survey estimates the number of green jobs for a NAICS industry based on the green activity found at individual establishments classified within the industry. The methodology does not simply designate an industry as “green” and count all jobs in that industry as GGS jobs, since establishments in the industry may also produce goods and services that are not considered green.

The BLS “output” definition of GGS employment does not include workers from all industries; instead, it focuses on employment within 333 of the approximately 1,192 detailed industries that exist under the NAICS classification. These 333 industries, the GGS scope, were selected by BLS after consultations with industry groups, government agencies, stakeholders, and the public helped identify industries that potentially provide green goods or services. These 333 industries were identified to provide goods and services that directly benefit the environment or conserve natural resources. Only the employment directly associated with the production of green goods and services within these selected industries are considered GGS jobs under the BLS definition.

Businesses and government establishments are assigned industry codes based on their primary activity. BLS recognizes that there may be some GGS employment in other businesses and government agencies in their secondary activities and that these are not counted within the GGS survey scope.

GGS Scope

The U.S. total annual average employment for 2010 was about 127,820,400 as measured by the Quarterly Census of Employment and Wages (QCEW) program. However, the GGS industries in their entirety contained about 25,513,300 workers, or about 20.0 percent of the nation's total employment.

Activities in the GGS scope included hydroelectric and nuclear energy production, as well as energy production from solar, wind, biomass, and other renewable resources. Industries within the GGS scope also produce green goods such as pollution control equipment, hybrid cars, recycled metals and paper, and organic farm products, among others. Green services provided by these selected industries include the operation of waste water treatment facilities, the operation of mass transit systems, the retail of used goods, the sale and trade of pollution control credits, and the enforcement of environmental regulations. These industrial activities occur both within the private sector as well as state, local and federal governments.

Well-recognized industry standards were used to distinguish between green products and others within the same establishment when appropriate. For example, manufacturing of LEED-rated materials and construction using LEED-rated materials are included. Appliances with an EnergyStar rating and products achieving the WaterSense rating are included.

About 120,000 business and government establishments were included in the survey. Each was asked to report the percent of revenue derived from green goods and services included in the definition. That percentage was multiplied by the employment level to derive the number of GGS jobs for that establishment. Non-profits, government units and business start ups without positive revenue were asked to supply a percent of employment. Thus, the employment figure includes workers of all occupations as long as they worked in the establishment with GGS employment or revenue. For example, a solar panel installation business might report that all of its revenue is included in the definition. In this case, all workers are counted, including installers, managers, secretaries, etc. Similarly, mass transit businesses reporting GGS revenue would include workers such as bus and subway drivers, maintenance and repair workers, managers and administrative personnel.

Sampling Methods

Sample Frame

The GGS uses the QCEW as its sampling frame. Private and Government (Federal, State, and Local) establishments are included on the frame, excluding any establishment with an average employment of zero over the past 12 months. The data for the QCEW come from State Unemployment Insurance files that are collected by individual State agencies. The QCEW includes several descriptive variables, such as name, address, monthly employment counts, industry classification, and geographic information for nearly all establishments in the United States.

Since it takes about one year for these data to be processed, the GGS frame for the 2011 second quarter (2011Q2) initial sample is comprised of 2010 second quarter (2010Q2) QCEW data. The 2010Q2 QCEW has over 8 million business establishments accounting for about 128 million employees. The GGS sample frame is restricted to the 333 in-scope industries and has approximately 2.1 million establishments accounting for about 25.5 million employees. To account for business openings in the third and fourth quarters of 2010, a small birth sample is selected from the 2010Q4 QCEW data.

About 13,000 in-scope establishments comprising approximately one million employees were pre-identified as being involved with some kind of green activity. These units were identified internally by BLS by use of the internet and an environmental database maintained by Environmental Business International (an environmental publishing, research and consulting company). The 13,000 establishments will be referred to as the Environmental Establishments frame and have special treatment during the GGS allocation and selection phases.

Sample Allocation

The GGS sample size is about 120,000 establishments:

- An initial sample of approximately 116,000 establishments is selected from the second quarter
- A birth sample of approximately 4,000 establishments is selected from the fourth quarter

The initial sample is allocated according to Table 1. Each type of frame unit has its own independent allocation.

Table 1: 2011 GGS Initial Allocation Breakouts

Type of Frame Unit	Sample Allocated
Private Establishments	94,500
Local Government Establishments	7,700
State Government Establishments	4,000
Federal Government Units	3,300
Environmental Establishments	6,500
Total	116,000

Private Allocation

The GGS private establishment allocation is stratified two-dimensionally by 1) State / 2-digit NAICS industries and 2) by 4 or 6 digit NAICS industries. These 4 or 6 digit industries will be called Allocation NAICS, or A_NAICS. For the most part the A_NAICS industries are at the 4-digit NAICS detail, however some industries that seemed to be highly environmental, such as 221119: Other electric power generation, (ex. 221119 – Other Electric Power Generation) were allocated to 6-digit detail.

The GGS private sample is first allocated by giving a minimum number of sample units to each State by 2-digit NAICS stratum. Second, a fixed number of sample units is allocated within each state using a power allocation, given in the formula below, and the larger of the two allocations is selected. This ensures that each State /2-digit NAICS has adequate sample representation for estimation.

$$n_h = n_s \frac{\sqrt{X_h}}{\sum_{h \in s} \sqrt{X_h}}$$

where

n_h = Amount of sample allocated to stratum h (State by 2-digit NAICS)

n_s = State sample size, initially equal to 1,000

X_h = Number of employees in stratum h

After the State level of allocation, the remaining sample is allocated nationally to A_NAICS industry strata, using a power allocation. The national sample size is increased and reconciled with the State sample in an iterative procedure until the total private allocation is close to the target sample size of 94,500.

Government Allocations

The sample units for Local, State, and Federal establishments use the same allocation methodology as the private sample allocation. The target sample sizes for each of these separate allocations are given in Table A.

Environmental Establishments Allocation

The Environmental Establishments frame includes establishments in the private and government sectors. The frame is stratified by 6-digit NAICS industry and establishment size. Generally, establishments from the frame were given a higher probability of inclusion in the sample. Approximately 6,500 establishments were allocated then selected from the Environmental Establishments frame.

GGS Sample Selection

The Private and Government samples are selected using a modified probability proportional to size (PPS) method where the size for an establishment is defined as:

$$size_i = \begin{cases} 10 & \text{if } X_i \leq 10 \\ X_i & \text{if } X_i > 10 \end{cases} \quad \text{where } X_i \text{ is frame employment for establishment } i$$

This type of sampling is sometimes referred to as PPZ sampling. The smallest establishments are treated differently because of the assumption that they have the potential for very large relative employment shifts between the time period of the QCEW data on the frame and when the establishment is collected. By raising the size of the smallest establishments, the selection probabilities are raised, reducing and stabilizing the sampling weights. A pure PPS sampling approach would create the potential for the smallest units to have very large sampling weights, which could create instability among small establishments with a high level of relative employment change.

Another modification to the PPS design arises from the two-dimensional survey design. Selection probabilities are reconciled between National/detailed industry and State/2-digit industry in such a way that an establishment's maximum selection probability from the two-dimensional allocation is assigned as its final probability of selection.

The Environmental Establishments sample is selected using simple random sampling within each 6-digit NAICS by size class stratum. Since the sample is allocated at a higher rate as the size class increases, there is an implicit probability proportional to size selection scheme.

GGS Birth Sample

A fourth quarter birth sample of about 4,000 establishments is selected to represent the newly-formed establishments that become in business or in-scope between the second quarter and the fourth quarter in 2010. Any establishment in the 2010Q4 sample frame that does not match to the 2010Q2 sample frame is considered a birth. The birth sample is allocated at the same rate as the initial sample for each of the five different allocations.

Sampling Weights

Each sampled establishment has a known probability of selection. The inverse of the probability of selection is called the sampling weight.

Estimation Methods

A Horvitz-Thompson (HT) estimator is used to estimate GGS employment. Establishment employment is updated from the Quarterly Census of Employment and Wages (QCEW) files relevant to the reference period of this release. Proportion of an establishment's total revenue associated with producing green goods or providing green goods is used to estimate the number of GGS employees contributed by the establishment. Each establishment has a sampling weight, which is the inverse of its probability of selection, that is multiplied by its number of green employees in the HT estimator. Weights are used since a sample of establishments, rather than the whole business population, is included in the survey.

The estimation levels of the GGS survey are

- National, total and private ownership
- 170 A_ NAICS industries, private
- Statewide, total and private
- State by 2-digit NAICS sector, private
- Renewable energy source, private (see Renewable Energy subsection)

In the estimation formulas, estimation cell ***h*** refers to any particular estimation level covered in the above list.

For establishment i in estimation cell h , let:

fw_i = Final weight¹ of i

e_i = 12-month average QCEW employment of i

p_i = Reported proportion of revenue² from GGS goods and services from i

\widehat{GE}_h = Estimated GGS employment in h

GGS employment for estimation cell h is calculated using the formula:

$$\widehat{GE}_h = \sum_{i \in h} (fw_i * e_i * p_i)$$

¹For details about the final weight, see Nonresponse and Benchmarking subsections.

²For non-revenue establishments, p_i is the reported proportion of employees involved with the production of green goods and services.

For calculation of GGS proportion of employment, \widehat{GE}_h is divided by the 12-month average of total QCEW employment for that estimation cell and converted into a percentage. Instead of being limited to the 333 NAICS industries classified in the GGS scope, all QCEW employment within that cell is included in the denominator. Thus, each GGS percentage estimate is relative to total employment in that cell. Out-of-scope industries are assumed to have zero GGS employment.

For estimation cell h , let

$QCEW_h$ = 12-month average QCEW employment in h

\widehat{GP}_h = Estimated GGS percentage of total employment in estimation cell h .

GGS percentage for estimation cell h is calculated using the formula:

$$\widehat{GP}_h = 100 * \left(\frac{\widehat{GE}_h}{QCEW_h} \right)$$

Nonresponse

For a variety of reasons, some establishments sampled by the GGS program either fail to respond or fail to provide complete, useable information on the returned survey form. Both are considered nonrespondents in the formulas below.

When calculating \widehat{GE}_h using the Horvitz-Thompson estimator, initial sample weights must be modified by a nonresponse adjustment factor to produce unbiased estimates of the true level of GGS employment. Nonresponse adjustment factors account for those sampled establishments that did not provide useable response information for inclusion in the estimation procedures.

The GGS survey produces estimates at both the National and State estimation levels. To accommodate the two-dimensional nature of the estimation levels, iterative, two-dimensional nonresponse adjustment factors are calculated for each respondent. Establishment size class is also incorporated due to the tendency of differential response rates among establishments of varying employment sizes. Size class is determined by maximum 12-month QCEW employment. In some industries, size class must be collapsed if there are too few respondents within a given size class.

There are two levels of nonresponse adjustment:

- National: 6-digit NAICS x size class

- State: State x 2-digit NAICS

The nonresponse adjustment procedure is iterated so that successive weighting adjustments converge to 1.00.

For establishment i in national adjustment cell h and State adjustment cell j , let

e_i = 12-month average QCEW employment of unit i

w_i = Sampling weight of unit i in nonresponse adjustment cells h and j

$w_{ih,k}$ = Nonresponse-adjusted weight of unit i in cell h , after the k^{th} first stage adjustment

$w_{ij,k}$ = Nonresponse-adjusted weight of unit i in cell j , after the k^{th} second stage adjustment

$NRAF_{ih,k}$ = Nonresponse adjustment factor of unit i in adjustment cell h after the first stage of iteration k

$NRAF_{ij,k}$ = Nonresponse adjustment factor of unit i in adjustment cell j after the second stage of iteration k

$FNRAF_i$ = Final nonresponse adjustment factor of unit i

S_h = Number of viable establishments in h (sampled establishments, excluding out of business and out of scope establishments)

R_h = Number of usable respondents in h , defined as units with Response Code 90 or 91

S_j = Viable establishments in j

R_j = Usable respondents in j , defined as units with Response Code 90 or 91

First Iteration

For the first iteration ($k = 1$), calculate:

$$NRAF_{ih,1} = \frac{\sum_{i \in S_h} w_i * e_i}{\sum_{i \in R_h} w_i * e_i}$$

$$w_{ih,1} = NRAF_{ih,1} * w_i$$

$$NRAF_{ij,1} = \frac{\sum_{i \in S_j} w_{ih,1} * e_i}{\sum_{i \in R_j} w_{ih,1} * e_i}$$

$$w_{ij,1} = NRAF_{ij,1} * w_{ih,1}$$

For the second and third iterations ($k = 2, 3$), calculate:

$$NRAF_{ih,k} = \frac{\sum_{i \in S_h} w_{ij,k-1} * e_i}{\sum_{i \in R_h} w_{ij,k-1} * e_i}$$

$$w_{ih,k} = NRAF_{ih,k} * w_{ij,k-1}$$

$$NRAF_{ij,k} = \frac{\sum_{i \in S_j} w_{ih,k} * e_i}{\sum_{i \in R_j} w_{ih,k} * e_i}$$

$$w_{ij,k} = \text{NRAF}_{ij,k} * w_{ih,k}$$

$$\text{FNRAF}_i = \frac{w_{ij,k}}{w_i}$$

Benchmarking

Due to differences between the date of initial sample selection and the reference period of the estimates, the GGS program updates the employment of each establishment to its 12-month average over the estimation reference period. Specifically, the initial sample is selected from the 2nd quarter QCEW files, while the reference period represents the calendar year ending in the 4th quarter. Establishments are linked to the 4th quarter, and their employment is updated with the new 12-month average.

As a result of this procedure, the additivity that the nonresponse procedure guarantees is lost because of new levels of QCEW employment. The GGS program calculates a benchmark factor to ensure that the estimates are additive. This procedure is similar to the nonresponse procedure with two primary differences:

1. National benchmarking occurs at the national estimation level, rather than the 6-digit industry x size class level.
2. QCEW employment totals are used in the numerator of the adjustment formulas, rather than weighted sampled employment.

The benchmarking procedure is iterated so that successive weighting adjustments converge to 1.00.

e_i = Average 2010 QCEW employment for establishment i

w_i = Sampling weight of unit i in estimation cells h and j

FNRAF_i = Final nonresponse adjustment factor of unit i

$aw_i = w_i * \text{FNRAF}_i$ = Nonresponse adjusted weight of unit i

$w_{ih,k}$ = Benchmark weight of unit i in cell h , after the k^{th} first stage adjustment

$w_{ij,k}$ = Benchmark weight of unit i in cell j , after the k^{th} second stage adjustment

$\text{BMF}_{ih,k}$ = Benchmark factor of unit i in adjustment cell h after the first stage of iteration k , where $k \leq 3$

$\text{BMF}_{ij,k}$ = Benchmark factor of unit i in adjustment cell j after the second stage of iteration k

FBMF_i = Final benchmark factor of unit i

N_h = Sample frame establishments in h

R_h = Usable respondents in h

N_j = Number of sample frame establishments in j

R_j = Usable respondents in j

First Iteration

For the first iteration ($k = 1$), calculate:

$$BMF_{ih,1} = \frac{\sum_{t \in N_h} e_t}{\sum_{t \in R_h} aw_t * e_t}$$

$$bw_{ih,1} = BMF_{ih,1} * aw_i$$

$$BMF_{ij,1} = \frac{\sum_{t \in N_j} e_t}{\sum_{t \in R_j} bw_{ih,1} * e_t}$$

$$bw_{ij,1} = BMF_{ij,1} * bw_{ih,1}$$

For the second and third iterations ($k = 2, 3$), calculate:

$$BMF_{ih,k} = \frac{\sum_{t \in N_h} e_t}{\sum_{t \in R_h} bw_{ij,k-1} * e_t}$$

$$bw_{ih,k} = BMF_{ih,k} * bw_{ij,k-1}$$

$$BMF_{ij,k} = \frac{\sum_{t \in N_j} e_t}{\sum_{t \in R_j} bw_{ih,k} * e_t}$$

$$bw_{ij,k} = BMF_{ij,k} * bw_{ih,k}$$

Calculate the final benchmark factor and final weight of establishment i :

$$FBMF_i = \frac{bw_{ij,k}}{aw_i}$$

$$fw_i = w_i * FNRAF_i * FBMF_i$$

Renewable Energy

Estimation levels in this release are in accordance with the 2007 NAICS industry codes. Establishments in the QCEW files were reclassified beginning in the first quarter of 2011 with the new 2012 NAICS industry codes. An industry reclassification of particular relevance to the measurement objective of the GGS survey is the breakout of industry 221119: Other electric power generation. The chart below displays the NAICS changes resulting from the reclassification.

2007 NAICS Industry Code	2012 NAICS Industry Code
221119: Other electric power generation	221114: Solar electric power generation
	221115: Wind electric power generation
	221116: Geothermal electric power generation
	221117: Biomass electric power generation
	221118: Other electric power generation

When the GGS sample was selected in the 2010Q2 sampling frame, the 2012 NAICS codes were not available to use for sample stratification. However, the GGS survey form for this industry asked for detailed response information for each of these types of other electric power generation. Discrepancies between the proportions of QCEW establishments in the breakout industries and the proportions of GGS respondents reporting the five energy

types were observed. To account for these differential response rates, special estimation techniques are employed to calculate estimates of Solar, Wind, Geothermal, Biomass, and Other electric power generation.

At the time of this release, the first two quarters of 2011 QCEW data were available with reclassified industry codes. Establishment responses to the GGS survey were compared to their classification in the 2011 QCEW files to ensure a good match between GGS respondents and their corresponding QCEW classification. Since a very high match was observed, the distribution of the number of original 221119 establishments that were broken out into the five new 2012 classifications was used to calculate post-stratification weights for estimation of the new categories of other electric power generation.

The distribution of establishments was calculated as the proportion, pe_k , of private establishments within a breakout industry k relative to the total number of private establishments in the five breakout industries.

Let

N_k = number of establishments in the 2011 QCEW classified within industry k , and

n_k = number of GGS respondents reporting a majority of revenue from the electric power generation type associated with k , where

$$k = \begin{cases} 1 & \text{for industry 221114} \\ 2 & \text{for industry 221115} \\ 3 & \text{for industry 221116} \\ 4 & \text{for industry 221117} \\ 5 & \text{for industry 221118} \end{cases}$$

P_k = proportion of establishments in the 2011 QCEW classified within industry k , relative to total number of establishments within industries 221114 – 221118, and

p_k = proportion of GGS respondents reporting a majority of revenue from the electric power generation type associated with k , relative to total number of respondents within industry 221119, calculated as

$$P_k = \frac{N_k}{\sum_{k=1}^5 N_k}$$

$$p_k = \frac{n_k}{\sum_{k=1}^5 n_k}$$

Post-stratification weight adjustments, pw_k , were calculated to generate green energy breakout estimates, gb_k , for the five types of other electric power generation. The post-stratification weights use auxiliary QCEW population information to correct for response bias existing in the renewable energy estimates.

$$pw_k = \frac{P_k}{p_k}$$

$$\widehat{gb}_k = \sum_{i \in k} [(aw_i * e_i * p_i) pw_k]$$

A raking factor, rf_h , was applied so the sum of the raked renewable energy breakout estimates, gb_k^* , equals the sum of the 221119 GGS employment estimate.

$$rf_h = \frac{GE_h}{\sum_{k=1}^5 gb_k}, \text{ where } h = \text{industry 221119}$$

$$\widehat{gb}_k^* = \widehat{gb}_k * rf_h$$

Reliability

Estimates developed from a sample will differ from the results of a census. An estimate based on a sample survey is subject to two types of error—sampling and nonsampling error. An estimate based on a census is only subject to nonsampling error.

Nonsampling error

This type of error is attributable to several causes, such as: errors in the sampling frame; an inability to obtain information for all establishments in the sample; differences in respondents' interpretation of a survey question; an inability or unwillingness of the respondents to provide correct information; errors made in recording, coding, or processing the data; and errors made in imputing values for missing data. Explicit measures of the effects of nonsampling error are not available.

Sampling errors

When a sample, rather than an entire population, is surveyed, estimates differ from the true population values that they represent. This difference, or sampling error, occurs by chance, and its variability is measured by the variance of the estimate or the standard error of the estimate (square root of the variance). The relative standard error is the ratio of the standard error to the estimate itself.

Estimates of the sampling error for the GGS employment estimates allow data users to determine if those statistics are reliable enough for their needs. Only a probability-based sample can be used to calculate estimates of sampling error. The formulas used to estimate GGS variances are adaptations of formulas appropriate for the survey design used.

The particular sample used in this survey is one of a large number of many possible samples of the same size that could have been selected using the same sample design. Sample estimates from a given design are said to be unbiased when an average of the estimates from all possible samples yields the true population value. In this case, the sample estimate and its standard error can be used to construct confidence intervals, or ranges of values that include the true population value with known probabilities. To illustrate, if the process of selecting a sample from the population were repeated many times, if each sample were surveyed under essentially the same unbiased conditions, and if an estimate and a suitable estimate of its standard error were made from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below to one standard error above the estimate would include the true population value. This interval is called a 68-percent confidence interval.
2. Approximately 90 percent of the intervals from 1.645 standard errors below to 1.645 standard errors above the estimate would include the true population value. This interval is called a 90-percent confidence interval.
3. Approximately 95 percent of the intervals from 1.96 standard errors below to 1.96 standard errors above the estimate would include the true population value. This interval is called the 95-percent confidence interval.
4. Almost all (99.7 percent) of the intervals from 3 standard errors below to 3 standard errors above the estimate would include the true population value.

Sample Variance

The GGS survey used a modified Balanced Repeated Replication (BRR) method to estimate sample variance. Prior to nonresponse adjustment and benchmarking, this replication technique splits the sample within an estimation cell into halves. The establishments within these half-samples have their sampling weights multiplied by 0.5 or 1.5, depending on the values of an orthogonal Hadamard matrix. All GGS estimates except Renewable Energy are then calculated with the modified sampling weights substituted for the original sampling weights described in the previous sections. Each replicate produces different half-samples and different estimates. The procedure is repeated γ times, and the resulting BRR estimates are used to estimate sample variances. Since the procedure is not applied to Renewable Energy breakout estimates, no sample variances are calculated for those estimates.

Estimated sample variance and standard error of the GGS employment estimate \widehat{GE}_h

$$\hat{V}(\hat{GE}_h) = \frac{1}{\gamma * (1 - \alpha)^2} \sum_{i=1}^{\gamma} (\hat{GE}_{h,\gamma}^{BRR} - \hat{GE}_h)^2$$

$$\hat{S}(\hat{GE}_h) = \sqrt{\hat{V}(\hat{GE}_h)}$$

Estimated sample variance and standard error of the GGS percentage estimate \hat{GP}_h

$$\hat{V}(\hat{GP}_h) = \left(\frac{100}{QCEW_h} \right)^2 \hat{V}(\hat{GE}_h)$$

$$\hat{S}(\hat{GP}_h) = \sqrt{\hat{V}(\hat{GP}_h)}$$

where

$\hat{GE}_{h,\gamma}^{BRR} = \gamma^{th}$ BRR estimate of GGS employment GE_h

\hat{GE}_h = Sample-based estimate of GGS employment GE_h

$\gamma = 128$ = Number of BRR replicates

$\alpha = 0.5$

Estimates of sample variance are useful to measure the precision of the GGS employment estimates, but they do have limitations. Generally, variances reflect the error component of the estimates that arise from surveying only a subset of the population, rather than the whole population. They do not reflect nonsampling error, such as response errors and nonresponse bias.

BLS-standard 90% confidence intervals for GGS employment may be constructed as:

$$\hat{GE}_h \pm 1.645 * \hat{S}(\hat{GE}_h)$$

If there was no nonsampling error in the survey, 90% of intervals constructed in this fashion, from all possible samples that could be selected, would contain the true population level of GGS employment for that estimation cell.